NASA TECHNICAL MEMORANDUM

NASA TM X-53357

November 4, 1965

NASA TM X-53357

N66-14912 (ACCESSION NUMBER) (THRU) 1

20 (PAGES) (CODE)

(NASA CR OR TMX OR AD NUMBER) 19 (CATEGORY)

FACILITY FORM 602

SIMULATION STUDY OF THE AMOUNT OF SENSITIVITY TEST DATA REQUIRED TO REJECT THE HYPOTHESIS OF NORMALITY WHEN THE SAMPLE POPULATION IS NONNORMAL

by J. B. GAYLE AND C. L. HOPKINS
Propulsion and Vehicle Engineering Laboratory

GPO PRICE $

CFSTI PRICE(S) $

Hard copy (HC) 1.00

Microfiche (MF) .50

ff 653 July 65

NASA

George C. Marshall Space Flight Center, Huntsville, Alabama

TECHNICAL MEMORANDUM X-53357

SIMULATION STUDY OF THE AMOUNT OF SENSITIVITY TEST DATA REQUIRED TO REJECT THE HYPOTHESIS OF NORMALITY WHEN THE SAMPLE POPULATION IS NONNORMAL

By J. B. Gayle and C. L. Hopkins

George C. Marshall Space Flight Center
Huntsville, Alabama

ABSTRACT

14912

Computer simulation techniques were used to study the number of sensitivity tests which are required to reject the hypothesis of a normally distributed sample population when the population actually was nonnormal. The results indicated that, even under the most favorable conditions, the number of tests required far exceed the number usually run in sensitivity type testing. This suggests that any assumption concerning the statistical nature of the distribution ordinarily will not be verified experimentally.

Author

NASA - GEORGE C. MARSHALL SPACE FLIGHT CENTER

TECHNICAL MEMORANDUM X-53357

SIMULATION STUDY OF THE AMOUNT OF SENSITIVITY TEST DATA REQUIRED TO REJECT THE HYPOTHESIS OF NORMALITY WHEN THE SAMPLE POPULATION IS NONNORMAL

By J. B. Gayle and C. L. Hopkins

PROPULSION AND VEHICLE ENGINEERING LABORATORY
RESEARCH AND DEVELOPMENT OPERATIONS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

TECHNICAL MEMORANDUM X-53357

SIMULATION STUDY OF THE AMOUNT OF SENSITIVITY TEST DATA REQUIRED TO REJECT THE HYPOTHESIS OF NORMALITY WHEN THE SAMPLE POPULATION IS NONNORMAL

SUMMARY

Computer simulation techniques were used to study the number of sensitivity tests which are required to reject the hypothesis of a normally distributed sample population when the population actually was nonnormal. The results indicated that, even under the most favorable conditions, the number of tests required far exceed the number usually run in sensitivity type testing. This suggests that any assumption concerning the statistical nature of the distribution ordinarily will not be verified experimentally.

When the stimulus level of particular interest to the experimenter is near the midpoint of the distribution, verification of the exact nature of the distribution is of little consequence. However, when the level of interest corresponds to a very high or very low probability of response, the results of this study indicate that a majority of the tests should be made at levels close to the level of interest, and nonparametric methods of statistical analysis should be used.

INTRODUCTION

When experimental data are analyzed, the assumption that the sample population is distributed normally is so generally accepted that frequently it is not stated. In many instances, this assumption does not introduce significant errors into the analysis, even when the distribution is markedly nonnormal. This is not always the case, however, and, in an earlier report (ref. 1), a computer simulation technique was used to demonstrate that, in the analysis of sensitivity (go/no go) test data, the use of statistical methods which assume the stimulus level versus reaction frequency relation to be represented by a cumulative normal distribution can introduce significant errors when the distribution actually is nonnormal. This indicates that either the assumption of normality should be verified or nonparametric statistical methods should be used for data of this type.

Although, generally, it is considered that the amount of experimental data needed to verify the assumption of normality is prohibitive, the actual amount depends on the characteristics of the particular distribution being studied; a general solution to this problem is not readily available. For sensitivity test data, however, the go/no go character of the data and the availability of an established procedure (ref. 2) for selecting the normal distribution giving the best fit for any particular set of data permit a relatively simple solution by computer simulation techniques. These same characteristics also permit an approximate analytical solution to the problem.

EXPERIMENTAL

Two sampling populations were used throughout this study. For the first, the stimulus level/frequency response relation was cumulative normal; for the second, this relation was linear. The overall process consisted of three parts.

The first consisted of selecting the test parameters, i.e., sampling population, stimulus levels, number of test results to be taken at each stimulus level, and the number of replicate experiments to be made. The selection of stimulus levels automatically determined the go/no go binomial probabilities used in the second stage of the process.

The second part of the process was the actual generation of test results. This was accomplished by means of a random number generator that was equipped with movable gates which could be adjusted to correspond to the binomial probabilities determined in the first part of the process.

The third part consisted of the analysis of the data. This included fitting a normal curve to the data by means of the Probit method of analysis and comparing the frequencies calculated from this fitted curve with the observed frequencies for the test data by means of the Chi square test for goodness of fit.

The overall process can be described by reference to FIG 1. In this figure, the straight line indicates the relation between the stimulus level and the frequency of responses in the population being sampled. The three stimulus levels selected for sampling are S_1, S_2, and S_3; the expected frequency of responses corresponding to S_1 is F_p. The frequencies that were observed when n samples were taken from each of the three stimulus levels are shown as X's, and that corresponding to stimulus S_1 is indicated as F_s. The cumulative normal curve shown was obtained by carrying out a standard Probit analysis using the three stimulus levels and corresponding sample frequencies. The frequency from this curve corresponding to stimulus level S_1 is F_n. The Chi square goodness of fit test is carried out by using the expected frequencies of response based on the fitted normal curve (F_n), the observed frequencies of response at each stimulus level (F_s), and the number of test results at each stimulus level (n). As a matter of convenience, no attempt was made to group terms having very low frequencies as is usually done in Chi square testing.

A check on the validity of the overall process was obtained by using a cumulative normal frequency distribution as the sampling population. For this population, significant Chi square values would not be expected, regardless of the number of samples, because the process consisted of fitting a normal curve to data drawn from a normally distributed population. The results were entirely consistent with this view; the observed Chi square values were unaffected by wide variations in the number of samples taken at each stimulus level and fell very close to those expected when no significant difference existed.

RESULTS

The selection of a sample population in which the relation between frequency of response and stimulus level was linear, i.e.,

$$F_p = S \qquad \text{(Eq. 1)}$$

was based on the fact that a linear relation of this type is the simplest mathematical relation to describe, requiring only one parameter, and that, in many instances, sensitivity test data appear to be linearly distributed. Also, a linear relation differs markedly from a cumulative normal distribution, especially at the higher (and lower) response levels; therefore, use of a linearly distributed sampling population should provide a conservative measure of the minimum number of tests needed to reject the hypothesis of normality when the population actually is nonnormal.

Several simulations were made to determine the effects of varying the number of samples taken at each stimulus level and also the number and location of stimulus levels. The number of samples at each level was varied from 20 to either 2,500 or 4,000 in each case; the number of stimulus levels and their corresponding binomial probabilities were as follows:

Values in Body of Table are Probabilities of Response at Given Stimulus Levels

Level No.	Simulation Number 1	2	3	4	5
1	.30	.20	.10	.50	.50
2	.40	.30	.20	.80	.75
3	.50	.40	.30	.85	.85
4	.60	.50	.40	.98	.90
5	.70	.60	.50		.92
6					.95
7					.97
8					.98
Chi Square*	9.4	9.4	9.4	7.8	14.1

* Chi Square needed to reject hypothesis of normality with 95 percent confidence.

The results are presented graphically in FIG 2; each plotted point represents the average of the Chi square values for ten replicate experiments.

Results for simulation #1 show the extreme difficulty in obtaining sufficient data for rejecting the hypothesis of normality when the stimulus levels selected for testing are uniformly distributed just above and below the 50 percent response level. Thus, Chi square values generally fell between 1.8 and 4.3, with only a slight trend toward increasing Chi square values being evident for increasing values of n. In no instance did the Chi square values approach the value of 9.4, which is required to reject the hypothesis of normality at the 95 percent confidence level, even when a total of 20,000 tests was used, i.e., 4,000 at each of 5 stimulus levels. Results for simulations #2 and #3 indicate the effects of progressively shifting the stimulus levels at which the samples were taken so that they are no longer distributed uniformly about the midpoint of the distribution. The results indicate, as expected, that the number of samples required to reject the hypothesis of normality decreases with increasing displacement of the response levels toward either end of the distribution.

When all of the selected levels are distributed between the midpoint and 90 percent level of the distribution, as for simulation #3, the number of tests at each level required to reject the hypothesis of normality is approximately 1,200, which totals 6,000 tests at all levels. By selecting the response levels to take advantage of the basic differences between linear and normal distributions, the number of tests can be decreased still further. However, this requires some knowledge about the sample population which would not be available in an experimental situation. Even so, the number of tests required for rejecting the hypothesis of normality is large, being approximately 800 for each of the cases tested, 200 at each of 4 levels for simulation #4 and 100 at each of 8 levels for simulation #5.

A few runs were made to determine the advantage afforded by dividing the total number of tests unevenly among the different stimulus levels as follows:

Stimulus Level	1	2	3	4
Probability of Response	0.50	0.80	0.85	0.95
Percent of Total Tests				
Simulation #6	.10	20	30	40
Simulation #7	40	30	20	10

The results may be summarized as follows:

Total Number of Tests	Average Chi Square for Given Simulation	
	#6	#7
500	7.5	6.5
1000	12.1	8.2
2000	16.8	11.7

NOTE: A Chi square value of 7.8 is required to reject the hypothesis of normality at the 95 percent confidence level.

Inspection of the data indicates a definite dependence of the Chi square values on the mode of distribution of the total number of tests among the different stimulus levels. In one instance, the difference in Chi square values for the two modes of sample distribution amounted to approximately 50 percent of the smaller value.

The excellent agreement between the simulation results and the general trends which were expected suggested that an approximate analytical solution to the problem should be possible. Considering a single Chi square term for responses, we have

$$\frac{(F_n - F_s)^2}{F_n} \qquad \text{(Eq. 2)}$$

The denominator of this term is available from the fitted normal curve. The numerator may be considered as a variance term made up of two components. The first component is the displacement between the population frequency and the normal curve frequency; the second component is the binomial variance of the sample results about the population frequency and can be estimated from the response probability and the number of samples. Thus, it is evident that

$$\frac{(F_n - F_s)^2}{F_n} \approx \frac{(F_n - F_p)^2 + F_p\left(1 - \frac{F_p}{n}\right)}{F_n} \qquad \text{(Eq. 3)}$$

By adding similar terms for non-responses and combining the resulting values for the different stimulus levels, predictions of the expected Chi square values for the various test conditions were obtained. Figure 3 presents the test data for simulation #4 and also a line which indicates the values predicted from equation 3. The excellent agreement between the experimental and predicted values appears to confirm the validity of both approaches to the problem.

DISCUSSION AND CONCLUSIONS

In interpreting these results, the number of tests required to reject the hypothesis of normality when the sample population is known to be nonnormal is assumed to be the minimum number required to verify the normality of a population which actually is normal.

The results of this study confirm the generally accepted belief that the amount of data needed to verify the assumption of normality generally is prohibitive. Although it is possible that use of more efficient goodness of fit tests would permit some decrease in the amount of data required, the difference between the number of tests usually employed in sensitivity testing (less than 100 distributed among several stimulus levels) and those required to reject the hypothesis of normality under the conditions selected for testing (usually more than 1,000) is so great that even appreciable increases in efficiency would be of little value. Therefore, it is evident that any assumption as to the statistical nature (normal, log normal, etc.) of any particular sample population must ordinarily remain unverified. When the intent of the experimenter is to establish the mean value or 50 percent response level for some variable, this factor is of little consequence because even an erroneous assumption of normality generally will not introduce serious errors into the analysis (ref. 1). Moreover, for such cases, it appears reasonable to further emphasize the importance of the central portion of the distribution by concentrating the sampling levels in this area as is done in the Bruceton method of analysis (ref. 3). However, when the intent of the experimenter is to establish the stimulus level corresponding to some very high or very low frequency of response, the sampling levels should be concentrated near the level of interest and nonparametric statistical methods should be used for analyzing the data.

REFERENCES

1. Gayle, John B.: Computer Simulation Study of Bruceton and Probit Methods of Sensitivity Testing. Proceedings of the Ninth Conference on the Design of Experiments in Army Research Development and Testing. ARO-D Report 64-2, 1964, pp. 205-227.

2. Finney, D. J.: Probit Analysis. Second ed., Cambridge University Press, 1962.

3. Statistical Research Group, Princeton University: Statistical Analysis for a New Procedure in Sensitivity Experiments. OSRD 4040, July 1944.

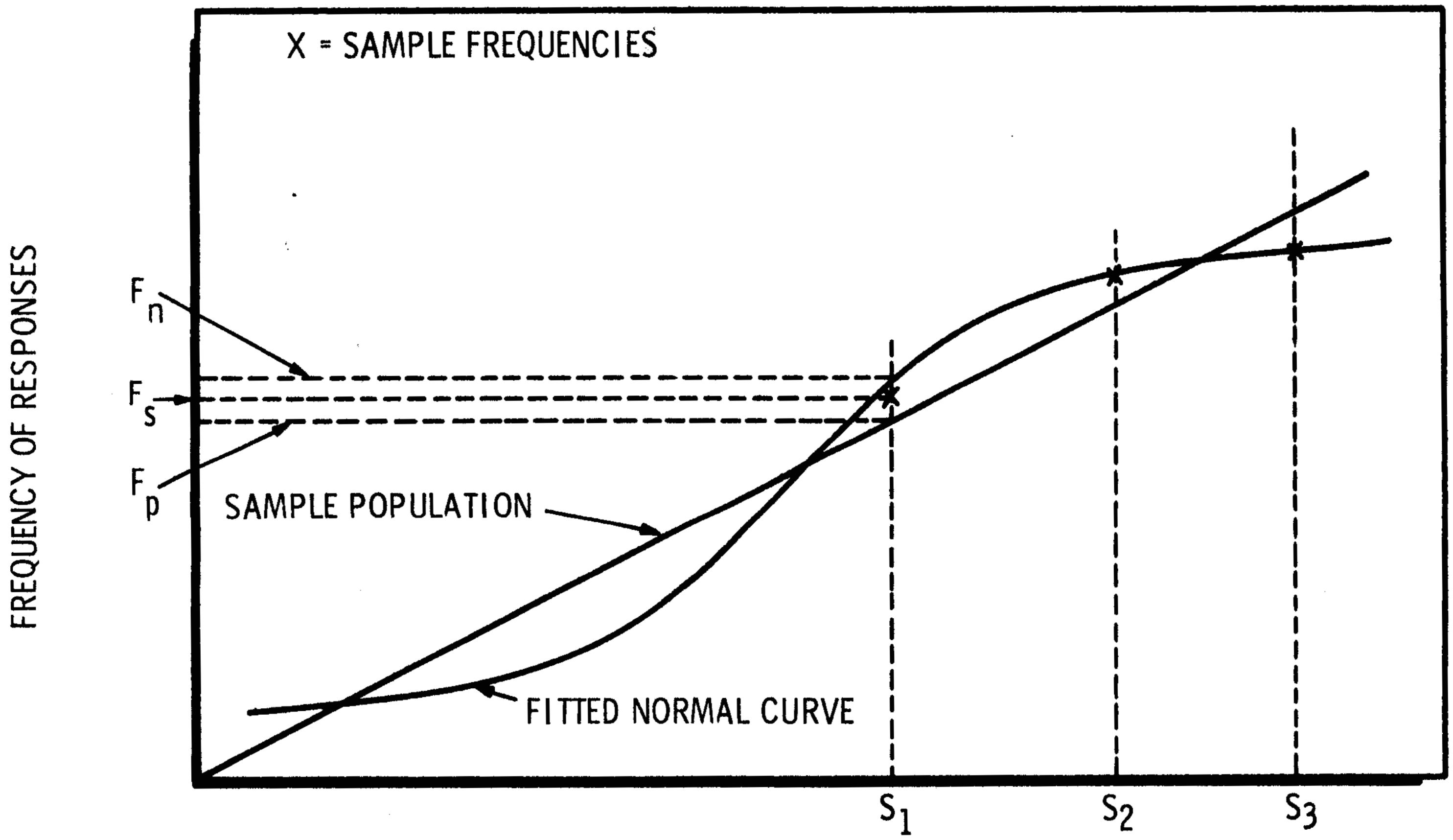

FIGURE 1. COMPARISON OF POPULATION, EXPERIMENTAL, AND FITTED NORMAL CURVE FREQUENCIES FOR A TYPICAL SIMULATION

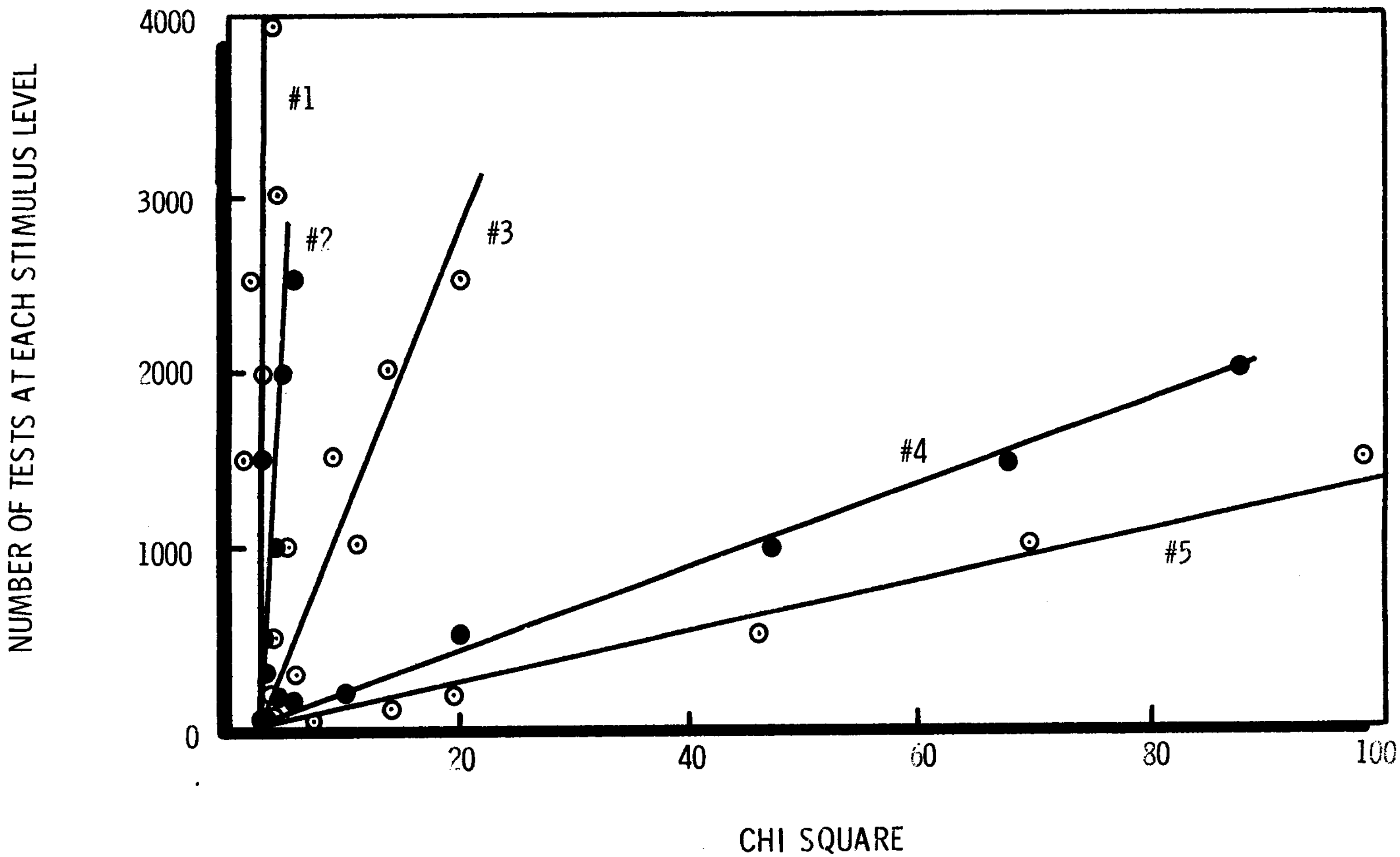

FIGURE 2. EFFECT OF NUMBER OF SAMPLES ON CHI SQUARE VALUES

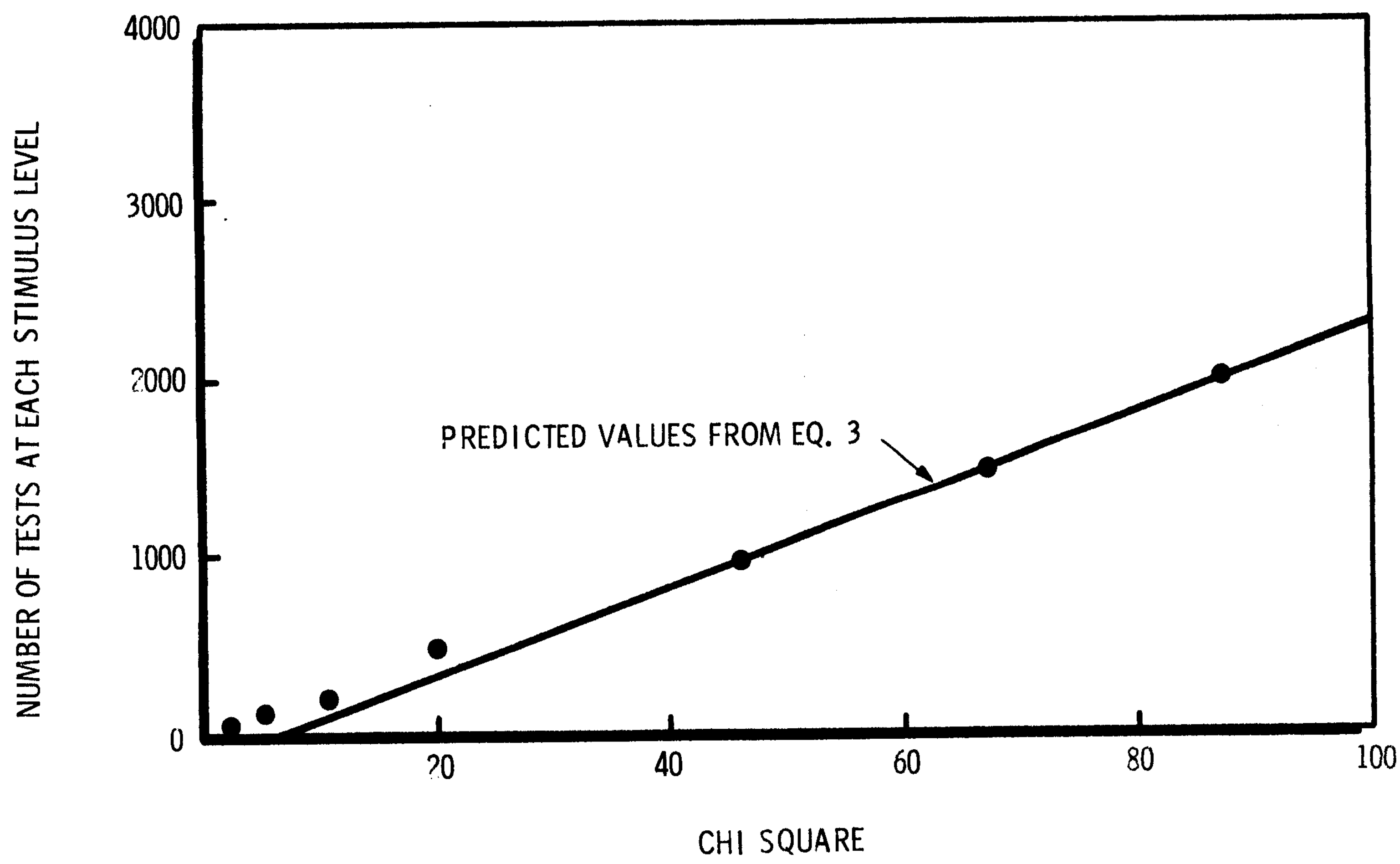

FIGURE 3. COMPARISON OF OBSERVED AND PREDICTED CHI SQUARE VALUES

November 4, 1965 APPROVAL NASA TM X-53357

SIMULATION STUDY OF THE AMOUNT OF SENSITIVITY TEST DATA REQUIRED TO REJECT THE HYPOTHESIS OF NORMALITY WHEN THE SAMPLE POPULATION IS NONNORMAL

By J. B. Gayle and C. L. Hopkins

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

W. A. Riehl

W. A. Riehl
Chief, Chemistry Branch

W. R. Lucas

W. R. Lucas
Chief, Materials Division

Erich A. Wellhausen

F. B. Cline
Director, Propulsion and Vehicle Engineering Laboratory

DISTRIBUTION NASA TM X-53357

R-ASTR-DIR	Dr. Haeussermann
R-ME-DIR	Mr. Kuers
R-P&VE-DIR	Mr. Cline
R-P&VE-DIR	Mr. Hellebrand
R-P&VE-M	Dr. Lucas (5)
R-P&VE-MC	Mr. Riehl (50)
R-P&VE-MM	Mr. Cataldo
R-P&VE-MN	Mr. Curry
R-P&VE-ME	Mr. Kingsbury
R-P&VE-P	Mr. Paul
R-P&VE-S	Mr. Kroll
R-P&VE-V	Mr. Aberg
R-QUAL-DIR	Mr. Grau (2)
R-P&VE-RT	Mr. Hofues
MS-H	Mr. Akens
MS-IP	Mr. Remer
MS-IL	Miss Robertson (8)
CC-P	Mr. Wofford
MS-T	Mr. Wiggins (5)
I-RM-M	Mr. Goldstein

NASA Scientific and Technical Information Facility (25)
P. O. Box 33
College Park, Maryland 20740

DIR

DEP-T